

THE SUP-NORM PROBLEM ON THE SIEGEL MODULAR SPACE OF RANK TWO

VALENTIN BLOMER AND ANKE POHL

ABSTRACT. Let F be a square integrable Maaß form on the Siegel upper half space \mathcal{H} of rank 2 for the Siegel modular group $\mathrm{Sp}_4(\mathbb{Z})$ with Laplace eigenvalue λ . If, in addition, F is a joint eigenfunction of the Hecke algebra and Ω is a compact set in $\mathrm{Sp}_4(\mathbb{Z}) \backslash \mathcal{H}$, we show the bound $\|F|_{\Omega}\|_{\infty} \ll_{\Omega} (1 + \lambda)^{1-\delta}$ for some global constant $\delta > 0$.

As an auxiliary result of independent interest we prove new uniform bounds for spherical functions on semisimple Lie groups.

1. INTRODUCTION

Given a Riemannian manifold X of finite volume, it is natural to ask for various properties of the L^2 -eigenfunctions of its Laplacian Δ . These eigenfunctions are of importance in a range of fields such as quantum chaos and mathematical physics, harmonic analysis and – if X has some connection to arithmetic – number theory. The relation to physics is provided by the fact that the Laplacian coincides, up to scaling, with the Schrödinger operator of a freely moving particle on X , and hence the L^2 -eigenfunctions of Δ are understood as bound states in physics. Hence typical questions ask about asymptotics (Weyl law), distribution and multiplicities of the eigenvalues of these L^2 -eigenfunctions, and the distribution of their masses along increasing sequences of eigenvalues. More refined questions include the asymptotic behaviour of Wigner distributions or microlocal lifts and their possible quantum limits (quantum ergodicity, quantum unique ergodicity, arithmetic quantum unique ergodicity); often entropy bounds play an important role in this context. For detailed surveys of recent results see, e.g. [Sar11, Zel10].

In this paper we study the problem of obtaining *pointwise* bounds in the spectral aspect for L^2 -normalized eigenfunctions on the Siegel modular space of rank 2, the quotient space of the Siegel upper half space of rank 2 by the Siegel modular group $\mathrm{Sp}_4(\mathbb{Z})$. Such upper bounds for sup-norms provide a measure for the equidistribution of mass of the respective eigenfunction since they estimate the extent to which the eigenfunctions may localize in small sets. The problem of bounding sup-norms is also closely related to the multiplicity problem: if V_{λ} denotes the eigenspace of the eigenvalue λ , then we have the inequality [Sar]

$$\dim V_{\lambda} \leq \mathrm{vol}(X) \sup_{\substack{\|F\|_2=1 \\ F \in V_{\lambda}}} \|F\|_{\infty}^2.$$

In particular, good bounds for sup-norms imply good bounds for multiplicities of eigenvalues, and conversely large eigenspaces prevent the possibility to obtain such bounds. A classical example (see e.g. [Far08, Iwa97]) is the sphere $X = S^2$; here all eigenvalues are of the form $\lambda = k(k+1)$ with $k \in \mathbb{N}_0$, and the corresponding eigenspaces are of dimension $2k+1$. One L^2 -normalized element in $V_{k(k+1)}$ is given by $F(\theta) := ((2k+1)/4\pi)^{1/2} p_k(\cos \theta)$ where θ is the azimuth angle and p_k is the k -th

2010 *Mathematics Subject Classification.* Primary 11F46, 11F72, 11H55, 43A90; Secondary 11D75, 11F60, 22E40.

Key words and phrases. Siegel Maaß forms, symplectic group, sup-norms, trace formula, Hecke operators, arithmetic subgroups, quadratic forms, diophantine approximation, spherical functions.

First author supported by the Volkswagen Foundation and a Starting Grant of the European Research Council. Second author supported by the Volkswagen Foundation.

Legendre polynomial. In particular,

$$\|F\|_\infty = \left(\frac{2k+1}{4\pi} \right)^{1/2} \asymp (1 + \lambda_F)^{1/4}.$$

In general, for any compact Riemannian manifold and L^2 -normalized Laplace eigenfunctions F , one has the bound [SS89, Hoe68, SZ02]

$$(1.1) \quad \|F\|_\infty \ll (1 + \lambda_F)^{(\dim X - 1)/4},$$

and the example of the n -sphere shows that this bound is sharp in general. However, if X carries additional symmetries, then it is reasonable to expect significantly stronger estimates. An important instance of the effect of additional symmetries has been established by Sarnak [Sar]: If X is a compact Riemannian locally symmetric space of rank r , then

$$(1.2) \quad \|F\|_\infty \ll (1 + \lambda_F)^{(\dim X - r)/4}$$

for joint eigenfunctions F with Laplace eigenvalue λ_F of the algebra of all differential operators that are invariant under the Riemannian isometry group of X [Sar]. For non-compact spaces, things are more complicated, and the lower bound for congruence quotients of $\mathrm{PGL}_n(\mathbb{R})$ in [BT, Theorem 1.2, Corollary 1.3] shows that (1.2) cannot serve as a global generic bound for non-compact locally symmetric spaces X , at least if the rank is large. However, when restricted to a fixed bounded subset $\Omega \subseteq X$, we have the corresponding generic bound

$$(1.3) \quad \|F|_\Omega\|_\infty \ll_\Omega (1 + \lambda_F)^{(\dim X - r)/4},$$

see [BT, (2.4)], even in the slightly stronger form $\|F|_\Omega\|_\infty \ll_\Omega (1 + \lambda_F)^{(\dim X - r)/4} \|F|_\Omega\|_2$ if Ω has non-empty interior.

Many classical examples of Riemannian locally symmetric spaces (in particular, all irreducible Riemannian symmetric spaces of higher rank) are arithmetic and hence enjoy additional symmetries given by the Hecke operators, a commutative family of normal operators. For the joint L^2 -eigenfunctions of the Hecke algebra and the invariant differential operators, multiplicity one results are known. Hence in absence of obvious obstructions one might hope to be able to further improve the bound (1.2) or (1.3) for these eigenfunctions.

The archetypical result of a power saving relative to the generic bound (1.2) is due to Iwaniec and Sarnak [IS95, Sar] in the situation $X = \Gamma \backslash \mathbb{H}$ where \mathbb{H} is the hyperbolic plane and $\Gamma \leq \mathrm{SL}_2(\mathbb{R})$ is a cocompact arithmetic lattice or $\mathrm{SL}_2(\mathbb{Z})$. For L^2 -normalized Hecke Maaß cusp forms F they proved the bound $\|F\|_\infty \ll (1 + \lambda_F)^{5/24 + \varepsilon}$.

Similar results have been obtained for the Hecke-Laplace eigenfunctions on the sphere and ellipsoids [Van97, BM13] and on congruence quotients of hyperbolic 3-space [BHM]. The underlying algebraic groups in these cases are $\mathrm{SL}_2(\mathbb{R}) = \mathrm{SO}_0(2, 1)$, $\mathrm{SO}(3)$ and $\mathrm{SL}_2(\mathbb{C}) = \mathrm{SO}_0(3, 1)$, all of which have real rank at most 1.

In this paper we consider for the first time a group of *real rank* 2, the symplectic group

$$G := \mathrm{Sp}_4(\mathbb{R}) \cong \mathrm{SO}_0(3, 2),$$

and an irreducible (arithmetic) lattice in G , namely the Siegel modular group

$$\Gamma := \mathrm{Sp}_4(\mathbb{Z}).$$

The Riemannian symmetric space of non-compact type associated to G is the Siegel upper half space \mathcal{H} of rank 2, and the Siegel modular space

$$X := \Gamma \backslash \mathcal{H}$$

is a non-compact 6-dimensional arithmetic Riemannian locally symmetric orbifold with one cusp. On X (or \mathcal{H}) we consider the (Γ -invariant) joint eigenfunctions in $L^2(X)$ of the algebra of G -invariant differential operators on \mathcal{H} and the Hecke algebra associated to Γ . These eigenfunctions are often called square integrable Hecke Siegel Maaß wave forms of genus (or degree) 2 for Γ , containing the

important subspace of Hecke Siegel Maaß cusp forms for Γ which are those forms that decay rapidly towards the cusp of X . For shortness we refer to all of these forms just as *joint eigenfunctions* in $L^2(X)$. For precise definitions and more details we refer to Sections 3 – 5 below and standard monographs on Siegel modular forms, e.g. [Fre83]. For the restriction of F to any compact set $\Omega \subseteq X$ we prove the following power saving of the bound (1.3):

mainthm

Theorem 1. *There exists $\delta > 0$ such that for any compact subset Ω of $\Gamma \backslash \mathcal{H}$ and any L^2 -normalized joint eigenfunction F in $L^2(\Gamma \backslash \mathcal{H})$ with Laplace eigenvalue λ_F we have*

$$\|F|_{\Omega}\|_{\infty} \ll_{\Omega} (1 + \lambda_F)^{1-\delta}.$$

Remarks. (1) Our proof provides an explicit value for δ ($> 10^{-6}$), but we have not optimized the numerical value.

(2) An inspection of the proof shows that we get better bounds when the Langlands parameters (μ_1, μ_2) of F approach the walls of the Weyl chambers. Using below (3.2) directly in (6.7) we obtain the bound

$$\ll ((1 + |\mu_1|)(1 + |\mu_2|)(1 + |\mu_1 + \mu_2|)(1 + |\mu_1 - \mu_2|))^{1/2-\delta}$$

in the situation of Theorem 1 (for some possibly different $\delta > 0$) which we slightly simplified as $(1 + |\mu_1|^2 + |\mu_2|^2)^{1-\delta}$. This is due to the fact that the Plancherel density drops close to the walls of the Weyl chambers, but on the other hand the behaviour of the spherical functions becomes more complicated as stationary points tend to blow up to non-trivial submanifolds. We deal with this problem in Theorem 2 below.

- (3) The statement of Theorem 1 includes the cuspidal and the residual spectrum, and in particular possible exceptional forms F where the Ramanujan conjecture fails to hold. Moreover, our result and proof is independent of any bounds towards the Ramanujan conjecture at finite places.
- (4) We do not investigate the behaviour in the cusp and treat only an arbitrary, but fixed compact piece of the manifold. Our principal method, based on a (pre-)trace formula, cannot distinguish between cuspidal and non-cuspidal constituents (in particular Eisenstein series) of the automorphic spectrum, as they are treated evenly. Therefore it is *a priori* clear that our sup-norm bounds must deteriorate as we approach the cusp. Of course, cusp forms are rapidly decaying towards the cusps. In the classical situation of genus 1 this can be quantified rather easily by the Fourier expansion, but it requires some highly non-trivial input such as precise uniform bounds for Bessel functions on the analytic side, and Rankin-Selberg theory combined with a famous and deep result of Hoffstein-Lockhart [HL94] on values of symmetric square L -functions at the edge of the critical strip on the arithmetic side. Such information is not available for genus 2. As mentioned earlier, the size of cusp forms towards the cusps is, in higher rank, a subtle issue, and it is not even obvious that (1.2) is true in our situation.
- (5) For completeness, it should be mentioned that there are various related problems which have been studied recently in the context of Hecke eigenforms on hyperbolic 2- and 3-manifolds: on the one hand one can study *lower* bounds which may arise from at least three different sources: (a) generic fluctuations that are slightly stronger than expected from the random wave model [Mil10]; (b) degenerate behaviour of special functions which leads to peaks high in the cusp [Sar, Tem, BT]; and (c) embedded submanifolds [RuSa94, Mil11] allowing eigenfunctions that are functorial lifts with non-generic behaviour. On the other hand, one can let the underlying space vary, and study the sup-norm of eigenfunctions on a sequence of covers X_N of X in terms of the volume of X_N [HT13, BM13, BHM].

We now describe some of the methods involved in the proof of Theorem 1. We view this as a prototype of a general higher rank situation, and we remark that the techniques can also be applied to treat related situations such as the case $G = GL(3)$ in a similar fashion. As pioneered in [IS95], the proof starts with an amplified pre-trace formula. We highlight at this point some of the novel ingredients in this paper.

The amplifier we use here is based on not only the standard Hecke operators for the lattice $\Gamma = \mathrm{Sp}_4(\mathbb{Z})$, but also involves those defined on individual double cosets. For the analysis of the amplifier, quite precise knowledge on their combinatorics in the Hecke algebra is required, which is rather unwieldy, but potentially useful in other situations. We refer to Sections 4 and 6 for details and the explicit construction of this amplifier which also implements some of the recent advances introduced in [Ven10] and [BHM].

The geometric side of the pre-trace formula yields a counting problem that reflects the algebraic structure of a maximal compact subgroup of the isometry group of the considered space. In our setup, the isometry group is $G = \mathrm{Sp}_4(\mathbb{R})$, and any fixed maximal compact subgroup K is isomorphic to $U(2)$ which is topologically (and almost algebraically) isomorphic to $S^1 \times S^3$. As one may therefore presume, the counting problem translates to twisted binary and quaternary quadratic forms. Roughly speaking, we need to bound the number of integral points in a δ -neighbourhood of the intersection of 4 particular quadrics in 8 variables. Our bound, Proposition 5 below, is essentially best possible for sufficiently small δ . We refer to Section 7, in particular to (7.2) and (7.5), for details. The counting techniques developed in this paper work in much greater generality.

Finally, we need uniform bounds for the inverse spherical transform of test functions localized at a given point in \mathfrak{a}^* (related to a spectral parameter and with large distance to the origin). This brings us to the second main result that we would like to highlight at this point. Bounds for inverse spherical transforms are often consequences of the decay of the spherical functions φ_λ on G/K for large parameter $\lambda \in \mathfrak{a}^*$ (see Sections 2 and 3 for the notation). Strong bounds for φ_λ have been obtained in particular by [DKV83, Theorem 11.1] and [Mar, Theorem 1.3], but neither of these bounds is sufficiently uniform for our purposes: the bound in [DKV83] requires the argument $\exp(H)$ of φ_λ to stay away from the identity in $G//K = K \backslash G/K = A/W$ by a fixed amount, and the bound in [Mar] requires the parameter λ to stay away from the walls of the Weyl chambers by a fixed amount. Although it might be possible to remove these assumptions, we proceed differently to prove the following uniform bound, which is also of independent interest:

sphericalbound

Theorem 2. *Let G be a real connected semisimple Lie group with finite center. Let $B \subseteq \mathfrak{a}$ be a bounded subset. Then for any $H \in B$ and any $\xi = \lambda + i\eta \in \mathfrak{a}^* + iC_\varrho$, the elementary spherical function φ_ξ of G with parameter ξ satisfies*

$$\varphi_\xi(\exp(H)) \ll_B (1 + \|\lambda\| \cdot \|H\|)^{-1/2}.$$

Unlike the bounds in [DKV83, Mar], the exponent is not best possible (we apply the stationary phase method only in one dimension), but it suffices for our application. The novelty here is the fact that the bound is completely uniform as λ approaches ∞ and/or walls of Weyl chambers and as $\exp(H)$ approaches the identity.

Acknowledgement: The authors would like to thank Jim Arthur, Farrell Brumley, Gergely Harcos and Ralf Schmidt for very helpful comments on various aspects of this work.

sec:bounds

2. SPHERICAL FUNCTIONS

2.1. Spherical functions and spherical transform. Let G be an arbitrary real connected semisimple Lie group with finite center, and let \mathfrak{g} denote its Lie algebra. Fix an Iwasawa decomposition $G = KAN$. Let \mathfrak{a} denote the Lie algebra of A , and let $H: G \rightarrow \mathfrak{a}$ be the Iwasawa projection defined by

$$g \in K \exp(H(g))N \quad \text{for all } g \in G.$$

Let \mathfrak{a}^* denote the dual of \mathfrak{a} , and $\mathfrak{a}_\mathbb{C}^*$ its complexification. Let Σ be the set of (restricted) roots of \mathfrak{g} , and fix a maximal subset Σ^+ of positive roots. Let ϱ denote the half-sum of the positive roots,

weighted with multiplicity:

$$\varrho = \sum_{\alpha \in \Sigma^+} \dim \mathfrak{g}_\alpha \cdot \alpha.$$

Here, \mathfrak{g}_α denotes the restricted root space of α . Let W denote the Weyl group of G . As usual, let $G//K$ denote the double quotient $K \backslash G / K$ and recall that it is isomorphic to A/W by Cartan decomposition. For later purposes we define the Cartan projection $C: G \rightarrow \mathfrak{a}/W$ via

$$(2.1) \quad g = k_1 \exp(C(g)) k_2 \quad \text{for appropriate } k_1, k_2 \in K.$$

The Killing form

$$\langle X, Y \rangle := \text{Tr}(\text{ad } X \circ \text{ad } Y)$$

defines an inner product on \mathfrak{a} , which carries over to \mathfrak{a}^* via root vectors. Thus, the induced inner product on \mathfrak{a}^* is given by $\langle \lambda, \mu \rangle = \langle H_\lambda, H_\mu \rangle$ for $\lambda, \mu \in \mathfrak{a}^*$, where $H_\lambda \in \mathfrak{a}$ is determined by $\lambda(H) = \langle H, H_\lambda \rangle$ for all $H \in \mathfrak{a}$. We denote the corresponding norms by $\|\cdot\|$. We note that the inner product on \mathfrak{a}^* extends in a straightforward way to a \mathbb{C} -bilinear symmetric form on $\mathfrak{a}_\mathbb{C}^*$, the complexification of \mathfrak{a}^* . We remark that the extension of $\langle \cdot, \cdot \rangle$ is not an inner product on $\mathfrak{a}_\mathbb{C}^*$. For $\lambda = \nu + i\mu$ with $\nu, \mu \in \mathfrak{a}^*$ we set $\|\lambda\|^2 := \|\nu\|^2 + \|\mu\|^2$, which coincides with $\langle \lambda, \lambda \rangle$ for $\lambda \in \mathfrak{a}^*$, but not for general $\lambda \in \mathfrak{a}_\mathbb{C}^*$.

From Harish-Chandra's monumental work it is known that the spherical functions on G are parametrized by $\mathfrak{a}_\mathbb{C}^*/W$. For any $\lambda \in \mathfrak{a}_\mathbb{C}^*$, the associated spherical function $\varphi_\lambda: G//K \rightarrow \mathbb{C}$ is given by

$$(2.2) \quad \varphi_\lambda(g) := \int_K e^{(\varrho + i\lambda)H(gk)} dk.$$

It is well-known that any two such functions φ_λ and φ_μ coincide if and only if $\lambda = w.\mu$ for some element $w \in W$, and φ_λ is bounded (by 1) if and only if $\lambda \in \mathfrak{a}^* + iC_\varrho$ ([HJ69]), where C_ϱ denotes the convex hull of the points $w.\varrho$, $w \in W$.

Let $C_c^\infty(G//K)$ denote the space of compactly supported bi- K -invariant smooth complex-valued functions. For $f \in C_c^\infty(G//K)$, its spherical transform is defined by

$$(2.3) \quad \tilde{f}(\lambda) := \int_G f(g) \varphi_{-\lambda}(g) dg.$$

We briefly recall the Paley–Wiener Theorem and the Harish-Chandra Inversion Formula for G . For more details see e.g. [Gan71] or [Hel08, Chap. IV]. For $R > 0$ let $\mathcal{H}^R(\mathfrak{a}_\mathbb{C}^*)$ denote the space of entire functions $f: \mathfrak{a}_\mathbb{C}^* \rightarrow \mathbb{C}$ that satisfy

$$f(\lambda) \ll_N (1 + \|\lambda\|)^{-N} e^{R|\text{Im } \lambda|} \quad \text{for all } \lambda \in \mathfrak{a}_\mathbb{C}^*$$

for each $N \in \mathbb{N}_0$. Let $\mathcal{H}_W^R(\mathfrak{a}_\mathbb{C}^*)$ denote the subspace of W -invariant functions in $\mathcal{H}^R(\mathfrak{a}_\mathbb{C}^*)$. Then

$$\mathcal{H}_W(\mathfrak{a}_\mathbb{C}^*) := \bigcup_{R>0} \mathcal{H}_W^R(\mathfrak{a}_\mathbb{C}^*)$$

is the space of Paley–Wiener functions. The inversion formula invokes the Harish-Chandra \mathbf{c} -function, which is the meromorphic function $\mathbf{c}: \mathfrak{a}_\mathbb{C}^* \rightarrow \mathbb{C}$ given by the Gindikin-Karpelevich product formula

$$\mathbf{c}(\lambda) = c_0 \cdot \prod_{\alpha \in \Sigma^+} \mathbf{c}_\alpha(\lambda).$$

For any $\alpha \in \Sigma^+$, the map \mathbf{c}_α is given by

$$(2.4) \quad \mathbf{c}_\alpha(\lambda) := \frac{2^{-i\lambda_\alpha} \Gamma(i\lambda_\alpha)}{\Gamma\left(\frac{i\lambda_\alpha}{2} + \frac{m_\alpha}{4} + \frac{1}{2}\right) \Gamma\left(\frac{i\lambda_\alpha}{2} + \frac{m_\alpha}{4}\right)}, \quad \lambda_\alpha := \frac{\langle \lambda, \alpha \rangle}{\langle \alpha, \alpha \rangle},$$

where $m_\alpha = \dim \mathfrak{g}_\alpha$. The constant c_0 is determined by $\mathbf{c}(-i\varrho) = 1$.

We endow $A/W \cong G//K$ with isometric metrics using the identification $\exp: \mathfrak{a} \rightarrow A$. We let $\overline{B_R^A(o)}$ denote the closed ball in A with radius R about the identity $o \in A/W$. Note that the walls of Weyl chambers are null sets with respect to $d\lambda/|\mathbf{c}(\lambda)|^2$.

paley

Proposition 3 (Paley–Wiener Theorem and Harish-Chandra Inversion Formula). *The spherical transform (2.3) is a bijection of $C_c^\infty(G//K)$ onto $\mathcal{H}_W(\mathfrak{a}_\mathbb{C}^*)$. For each $R > 0$, it restricts to a bijection of the space of functions in $C_c^\infty(G//K)$ with support in $\overline{B_R^A(o)}$ onto $\mathcal{H}_W^R(\mathfrak{a}_\mathbb{C}^*)$. Its inverse is given by*

$$f(g) = \frac{1}{|W|} \int_{\mathfrak{a}^*} \tilde{f}(\lambda) \varphi_\lambda(g) \frac{d\lambda}{|\mathbf{c}(\lambda)|^2}.$$

We see that bounds for the inverse spherical transform f can be obtained from bounds of φ_λ . We now prepare for the proof of Theorem 2. We let the elements of the real universal enveloping algebra $U(\mathfrak{k})$ act as differential operators on K . Thus, if $f: K \rightarrow \mathbb{C}$ is a smooth function and $Y = X_1 \dots X_d$ ($X_i \in \mathfrak{k}$) is a homogeneous tensor of degree d over \mathfrak{k} , then

$$f(k; Y) = \frac{d^r}{dt_1 \dots dt_d} \Big|_{t_1=\dots=t_d=0} f(k \exp(t_1 X_1) \dots \exp(t_d X_d)).$$

For the proof of Theorem 2 we take advantage of the following weak van der Corput lemma, which is easily proven by applying [Ste93, Proposition 5, pp. 342–343].

vdCorput

Lemma 4. *Let K be a compact Lie group with Lie algebra \mathfrak{k} and Haar measure dk . Let $\psi: K \rightarrow \mathbb{C}$ be a smooth function and $\phi: K \rightarrow \mathbb{R}$ be smooth and real-valued. Suppose that there exists a (finite) open cover $(U_j)_{j=1}^m$ of K , a constant $c > 0$ and a constant $d \in \mathbb{N}$ such that for each $j \in \{1, \dots, m\}$ there exists a homogeneous tensor Y_j over \mathfrak{k} of degree at most d with $|\phi(k; Y_j)| \geq c$ for all $k \in U_j$. Suppose further that there exists a constant $N > 0$ such that the C^{d+1} -norm of ϕ and the C^1 -norm of ψ are bounded by N . Then*

$$\left| \int_K \psi(k) e^{it\phi(k)} dk \right| \ll t^{-\frac{1}{d}} \quad (t \rightarrow \infty).$$

The implied constant depends on ϕ and ψ and only through c, d, N .

For the proof of Theorem 2 we consider the integral in (2.2) as an oscillatory integral, where the directions of the argument $a = \exp(H) \in A/W = G//K$ and λ are fixed and their length are the variables. The phase function is obviously linear in the length of λ , and we will linearize it in the length of H .

2.2. Proof of Theorem 2. As recalled above, $\|\varphi_\xi\|_\infty \leq 1$ for all $\xi \in \mathfrak{a}^* + iC_\varrho$.

To prove the $\phi_\xi(\exp(H)) \ll (\|\lambda\| \cdot \|H\|)^{-1/2}$ for H in a bounded set B , we first observe that [DKV83, Theorem 11.1] yields a stronger bound as long as H stays away from the identity. We can therefore assume without loss of generality that $\|H\| \leq C_0$ for a sufficiently small constant C_0 . We write $a = a_\delta = \exp(\delta X) \in \mathfrak{a}$ for some $0 < \delta < C_0$ and $\|X\| = 1$ and $\lambda = \lambda_\tau = \tau \langle \cdot, Y \rangle \in \mathfrak{a}^*$ for $\|Y\| = 1$. As a and λ vary, X and Y vary in a compact set. For $a \in \mathfrak{a}$ we set

$$F_{a,Y}: K \rightarrow \mathbb{R}, \quad F_{a,Y}(k) := \langle H(ak), Y \rangle,$$

and for $\eta \in \mathfrak{a}^*$ we let

$$\psi_\eta: A \times K \rightarrow \mathbb{R}, \quad \psi_\eta(a, k) := e^{(\varrho - \eta)H(ak)}.$$

As long as η is fixed (e.g. contained in C_ρ) and a is bounded, ψ and its derivatives are uniformly bounded. With this notation we have

notation

$$(2.5) \quad \varphi_{\lambda_\tau + i\eta}(a_\delta) = \int_K \psi_\eta(a_\delta, k) e^{i\tau F_{a_\delta, Y}(k)} dk.$$

To linearize the phase function $F_{a_\delta, Y}$ in the variable δ , let $f_{X,Y}: K \rightarrow \mathbb{R}$ be defined as

$$f_{X,Y}(k) := \langle X, \text{Ad}_k(Y) \rangle.$$

By [DKV83, Proposition 5.5],

$$\left. \frac{d}{d\delta} \right|_{\delta=0} F_{\exp(\delta X), Y}(k) = f_{X, Y}(k),$$

and hence

$$F_{a_\delta, Y}(k) = \delta f_{X, Y}(k) + O(\delta^2).$$

For the derivative and the Hessian of $F_{a_\delta, Y}$ this implies

$$F_{a_\delta, Y}(k; U) = \delta f_{X, Y}(k; U) + O(\delta^2), \quad F_{a_\delta, Y}(k; UV) = \delta f_{X, Y}(k; UV) + O(\delta^2)$$

for all $U, V \in \mathfrak{k}$. By [DKV83, Proposition 1.2], the critical set of $f_{X, Y}$ is $C_{X, Y} := K_X W K_Y$, where K_X denotes the centralizer of X in K . Let $D_{X, Y}$ denote the subset of K on which the Hessian of $f_{X, Y}$ vanishes. By [DKV83, Proposition 1.4, Corollary 1.5],

$$C_{X, Y} \cap D_{X, Y} = \emptyset.$$

Let U_C, U_D be disjoint open neighbourhoods in K of $C_{X, Y}, D_{X, Y}$, respectively. Fix a basis V_1, \dots, V_m of \mathfrak{k} (here, $m = \dim \mathfrak{k}$). Then, for some constant $C_1 > 0$,

$$\begin{aligned} \max_{1 \leq i \leq m} |f_{X, Y}(k; V_i)| &\geq C_1 & \text{for all } k \in U_C, \\ \max_{1 \leq i, j \leq m} |f_{X, Y}(k; V_i V_j)| &\geq C_1 & \text{for all } k \in U_D. \end{aligned}$$

By choosing C_0 sufficiently small, it follows that

$$\begin{aligned} \max_{1 \leq i \leq m} |F_{a_\delta, Y}(k; V_i)| &\geq \delta C_1 / 2 & \text{for all } k \in U_C, \\ \max_{1 \leq i, j \leq m} |F_{a_\delta, Y}(k; V_i V_j)| &\geq \delta C_1 / 2 & \text{for all } k \in U_D. \end{aligned}$$

We rewrite (2.5) as

$$\varphi_{\lambda_\tau + i\eta}(a_\delta) = \int_K \psi_\eta(a_\delta, k) \exp\left(i\tau\delta(\delta^{-1}F_{a_\delta, Y}(k))\right) dk.$$

By the above, K can be covered by finitely many open subsets on each of which a first or second order partial derivative of $\delta^{-1}F_{a_\delta, Y}$ is uniformly bounded from below, and all its partial derivatives up to order 3 are uniformly bounded from above. Thus, by Lemma 4,

$$\varphi_{\lambda_\tau + i\eta}(a_\delta) \ll (\tau\delta)^{-\frac{1}{2}} = (\|\lambda_\tau\| \cdot \|\log a_\delta\|)^{-\frac{1}{2}}.$$

□

3. THE SIEGEL UPPER HALF PLANE AND $\mathrm{Sp}_4(\mathbb{R})$

sec2

In the following sections we set up the scene for the proof of Theorem 1. From now on we write

$$G := \mathrm{Sp}_4(\mathbb{R}) = \{M \in \mathrm{GL}_4(\mathbb{R}) \mid MJM^\top = J\}$$

with $J := \begin{pmatrix} & & & I_2 \\ & & & \\ & & & \\ -I_2 & & & \end{pmatrix}$. A standard choice for an Iwasawa decomposition $G = KAN = NAK$ is

$$\begin{aligned} K &= \left\{ \begin{pmatrix} B & C \\ -C & B \end{pmatrix} \in O(4) \right\}, \\ A &= \left\{ \mathrm{diag}(e^{t_1}, e^{t_2}, e^{-t_1}, e^{-t_2}) \mid t_1, t_2 \in \mathbb{R} \right\} \end{aligned}$$

and

$$N = \left\{ \begin{pmatrix} B & C \\ & B^{-\top} \end{pmatrix} \mid BC^\top = CB^\top, B \text{ unit upper triangular} \right\}.$$

We note that $K \cong U(2)$ via

$$\begin{pmatrix} B & C \\ -C & B \end{pmatrix} \mapsto B + iC.$$

compact

(3.1)

The Lie algebra of A is

$$\mathfrak{a} := \{\text{diag}(t_1, t_2, -t_1, -t_2) \mid t_1, t_2 \in \mathbb{R}\}.$$

For $j = 1, 2$, we define $e_j \in \mathfrak{a}^*$ by

$$e_j(\text{diag}(d_1, d_2, -d_1, -d_2)) = d_j.$$

The 8 (restricted) roots of $(\mathfrak{g}, \mathfrak{a})$ are $\Sigma := \{\pm e_1 \pm e_2, \pm 2e_1, \pm 2e_2\}$ (a C_2 root system). One easily checks that the multiplicity of each root is 1. A standard choice of positive roots is

$$\Sigma^+ := \{e_1 \pm e_2, 2e_1, 2e_2\}.$$

From (2.4) one checks that

cexplicit

$$(3.2) \quad |\mathbf{c}(\lambda)|^{-2} = \left(\frac{\pi}{4}\right)^2 \gamma(\lambda_1)\gamma(\lambda_2)\gamma(\lambda_1 + \lambda_2)\gamma(\lambda_1 - \lambda_2) \ll \|\lambda\|^4, \quad \text{where } \gamma(x) := x \tanh(\pi x/2).$$

Let

$$\mathcal{H} := \{Z = X + iY \in \text{Mat}_2(\mathbb{C}) \mid Z = Z^\top \text{ and } Y > 0\}$$

denote the Siegel upper half space with Riemannian metric determined by the line element $ds^2 = \text{Tr}(dZ Y^{-1} d\bar{Z} Y^{-1})$. Here, $Y > 0$ means that the matrix Y is positive definite. The action of G on \mathcal{H} is given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot Z = (AZ + B)(CZ + D)^{-1},$$

which induces an isometry of manifolds and G -spaces between G/K and \mathcal{H} via $gK \mapsto g.iI_2$. One representative in G of a point $Z = X + iY \in \mathcal{H}$ is given by

point1

$$(3.3) \quad g = \begin{pmatrix} I_2 & X \\ & I_2 \end{pmatrix} \begin{pmatrix} V & \\ & V^{-1} \end{pmatrix} \in G$$

where V is the unique symmetric positive definite matrix satisfying $V^\top V = Y$.

Let $\mathcal{D}(\mathcal{H})$ denote the algebra of differential operators on \mathcal{H} which are invariant under the left action of G . This is a commutative algebra of rank 2 which contains the positive definite Laplace-Beltrami operator Δ on \mathcal{H} . The Harish-Chandra isomorphism [Hel84, Chapter II.5] establishes a bijection between $\mathcal{D}(\mathcal{H})$ and the algebra of Weyl group invariant polynomials in $\mathfrak{a}_{\mathbb{C}}^*$. In particular, the image of Δ is

laplace

$$(3.4) \quad \langle \varrho, \varrho \rangle + \langle \lambda, \lambda \rangle = \frac{5 + \lambda_1^2 + \lambda_2^2}{12} \in \mathbb{C}[\lambda_1, \lambda_2]$$

for $\lambda = \lambda_1 e_1 + \lambda_2 e_2 \in \mathfrak{a}_{\mathbb{C}}^*$.

4. HECKE OPERATORS

sec3

If \mathcal{M} is a set of matrices in

$$\text{GSp}_4^+(\mathbb{Q}) = \{M \in \text{GL}_4^+(\mathbb{Q}) \mid MJM^\top = rJ \text{ for some } r \in \mathbb{Q}^*\}$$

that is left- and right-invariant under $\Gamma = \text{Sp}_4(\mathbb{Z})$ and a finite union $\bigcup_j \Gamma M_j$ of left cosets (equivalently, \mathcal{M} is a finite union of double cosets), then \mathcal{M} defines the associated Hecke operator on functions $F : \Gamma \backslash \mathcal{H} \rightarrow \mathbb{C}$ by

$$T_{\mathcal{M}} : F \mapsto \sum_j F \left(\frac{1}{(\det M_j)^{1/4}} M_j \cdot \right).$$

This definition extends in an obvious way to the vector space of formal linear combinations of such sets \mathcal{M} , and we obtain the symplectic Hecke algebra \mathcal{H} , a commutative algebra of operators that commute with the elements in $\mathcal{D}(\mathcal{H})$ and that are Hermitian with respect to the inner product

inner

$$(4.1) \quad \langle F_1, F_2 \rangle := \int_{\Gamma \backslash \mathcal{H}} F_1(Z) \overline{F_2(Z)} \frac{dX dY}{(\det Y)^3}.$$

We refer to [Fre83, Section 4] or [AZ95, Section 3] for proofs of these facts, historical remarks, and an introduction to the theory of Hecke algebras. The composition of two such operators corresponds to the multiplication of two double cosets $\Gamma A \Gamma = \bigcup_j \Gamma A_j$, $\Gamma B \Gamma = \bigcup_k \Gamma B_k$:

$$T_{\Gamma A \Gamma} \circ T_{\Gamma B \Gamma} : F \mapsto \sum_{j,k} F \left(\frac{1}{(\det A_j B_k)^{1/4}} A_j B_k \cdot \right).$$

It is easy to see that

$$T_{\Gamma A \Gamma} \circ T_{\Gamma B \Gamma} = \sum_D c_D T_{\Gamma D \Gamma}$$

where D runs through a system of representatives of double cosets contained in $\Gamma A \Gamma B \Gamma$ and c_D is the number of pairs (j, k) such that $\Gamma D = \Gamma A_j B_k$. However, in explicit situations this formula is combinatorially cumbersome.

For $m \in \mathbb{N}$ let

$$S(m) := \{M \in \mathrm{GSp}_4^+(\mathbb{Z}) \mid M^\top J M = mJ\}.$$

Then the m -th Hecke operator is given by $T(m) := T_{S(m)}$. For $(m_1, m_2) = 1$ we have

$$T(m_1)T(m_2) = T(m_1 m_2).$$

For $r \in \mathbb{N}_0$, $0 \leq a \leq b \leq r/2$ and any prime p define

$$T_{a,b}^{(r)}(p) := T_{\Gamma \mathrm{diag}(p^a, p^b, p^{r-b}, p^{r-a}) \Gamma}.$$

Then $T(p^r)$ can be decomposed as a sum over Hecke operators on individual double cosets:

$$T(p^r) = \sum_{0 \leq a \leq b \leq r/2} T_{a,b}^{(r)}(p).$$

Note that for all $a \in \mathbb{N}_0$ we have

$$T_{a,a}^{(2a)}(p) = \mathrm{id}, \quad \text{hence} \quad T_{a,b}^{(r)}(p) = T_{0,b-a}^{(r-2a)}(p) T_{a,a}^{(2a)}(p) = T_{0,b-a}^{(r-2a)}(p).$$

For a function $F : \Gamma \backslash \mathcal{H} \rightarrow \mathbb{C}$ that is an eigenfunction of the Hecke algebra \mathcal{H} , we denote by $\lambda(m, F)$ and $\lambda_{a,b}^{(r)}(p, F)$ the Hecke eigenvalue of F with respect to $T(m)$ and $T_{a,b}^{(r)}(p)$, respectively. In order to construct an efficient amplifier for the pre-trace formula we need to understand the Hecke relations. The formal generating series for Hecke operators is given by (combine [Shi63, Theorem 2] with (4.6) and (4.8) below)

$$\sum_{r=0}^{\infty} T(p^r) X^r = \frac{1 - p^2 X^2}{1 - T(p)X + (T(p)^2 - T(p^2) - p^2)X^2 - p^3 T(p)X^3 + p^6 X^4}.$$

Comparing coefficients for $r = 4$, we conclude that

$$T(p^4) = (p^2 + 2p^3)T(p)^2 - T(p)^4 + p^2 T(p^2) + T(p^2)T(p)^2 + T(p^2)^2 - p^6$$

so that $\lambda(p, F)$, $\lambda(p^2, F)$ and $\lambda(p^4, F)$ cannot be simultaneously small; quantitatively:

$$|\lambda(p, F)| + \frac{1}{p^{3/2}} |\lambda(p^2, F)| + \frac{1}{p^{9/2}} |\lambda(p^4, F)| \gg p^{3/2}.$$

(It has been proved in [SV, Lemma 5.2] that an inequality of this type exists in great generality.)

In order to analyze the amplifier in Section 6 below, we need an explicit decomposition of $T(p^r)^2$ for $r = 1, 2, 4$ into double cosets. The Hecke relation

$$T(p)^2 = T_{0,0}^{(2)}(p) + (p+1)T_{0,1}^{(2)}(p) + (p^3 + p^2 + p + 1)T_{1,1}^{(2)}(p)$$

(see e.g. [BvdGHZ08, p. 219]) is well-known. Unfortunately very little explicit is in the literature for higher powers, and the computations become indeed very involved. The only reference¹ we are aware of is [Kod67, p. 120] from which we quote the following for $r \geq 2$ (add the three columns of the table in the middle of the page and combine with (4.6)):

$$\begin{aligned}
 T(p^r)T(p^2) &= T_{0,0}^{(r+2)}(p) + (p+1)T_{0,1}^{(r+2)}(p) + (p^2+p+1) \sum_{b=2}^{(r+2)/2} T_{0,b}^{(r+2)}(p) \\
 &+ (p^3+p^2+p+1)T_{1,1}^{(r+2)}(p) + (p^4+2p^3+p^2+p+1) \sum_{b=1}^{r/2} T_{0,b}^{(r)}(p)T_{1,1}^{(2)}(p) \\
 &+ (p^6+p^5+2p^4+2p^3+p^2+p+1) \sum_{a=1}^{r/2} \sum_{b=0}^{(r-2a)/2} T_{0,b}^{(r-2a)}(p)T_{a+1,a+1}^{(2a+2)}(p).
 \end{aligned}
 \tag{4.9}$$

With $r = 2$ we obtain an exact formula for $T(p^2)^2$. To decompose $T(p^4)^2$ into a linear combination of Hecke operators $T_{a,b}^{(8)}(p)$ with $0 \leq a \leq b \leq 4$, we content ourselves with fairly crude upper bounds for the coefficients.

To this end we introduce the following notation: if

$$T_1 = \sum_{0 \leq a \leq b \leq r/2} \gamma_1(a, b) T_{a,b}^{(r)}(p), \quad T_2 = \sum_{0 \leq a \leq b \leq r/2} \gamma_2(a, b) T_{a,b}^{(r)}(p)$$

are two Hecke operators, we write $T_1 \leq T_2$ if $\gamma_1(a, b) \leq \gamma_2(a, b)$ for all a, b (note that the coefficients $\gamma_j(a, b)$ in such decompositions are unique). With this notation we obtain from (4.5) with $2r+2$ in place of r and (4.9) with $2r$ in place of r that

$$\begin{aligned}
 (4.10) \quad T(p^{2r+2}) &\leq T(p^{2r})T(p^2) \\
 &\leq 3p^2 \sum_{b \leq r+1} T_{0,b}^{(2r+2)}(p) + 6p^4 \sum_{b \leq r} T_{0,b}^{(2r)}(p)T_{1,1}^{(2)}(p) + 9p^6 \sum_{1 \leq a \leq r} \sum_{b \leq r-a} T_{0,b}^{(2r-2a)}(p)T_{a+1,a+1}^{(2a+2)}(p) \\
 &\leq 10 \sum_{s \leq r+1} p^{2r+4-2s} \sum_{b \leq s} T_{0,b}^{(2s)}(p)T_{r+1-s, r+1-s}^{(2r+2-2s)}(p).
 \end{aligned}$$

This remains trivially true for $r = 0$. With $r = 1$, we conclude $T(p^4) \leq T(p^2)^2$, and hence $T(p^4)^2 \leq T(p^4)T(p^2)T(p^2)$. Applying the upper bound in (4.10) twice (with $r = 2$ and $r = s$) in connection with (4.5), we obtain

$$\begin{aligned}
 T(p^4)^2 &\leq T(p^4)T(p^2)T(p^2) \leq 10 \sum_{s \leq 3} p^{8-2s} \sum_{b \leq s} T_{0,b}^{(2s)}(p)T_{3-s, 3-s}^{(6-2s)}(p)T(p^2) \\
 &\leq 10 \sum_{s \leq 3} p^{8-2s} T(p^{2s})T_{3-s, 3-s}^{(6-2s)}(p)T(p^2) \\
 (4.11) \quad &\leq 100 \sum_{s \leq 3} \sum_{\tau \leq s+1} p^{12-2\tau} \sum_{b \leq \tau} T_{0,b}^{(2\tau)}(p)T_{4-\tau, 4-\tau}^{(8-2\tau)}(p) \\
 &\leq 400 \sum_{0 \leq \tau \leq 4} p^{12-2\tau} \sum_{b \leq \tau} T_{0,b}^{(2\tau)}(p)T_{4-\tau, 4-\tau}^{(8-2\tau)}(p).
 \end{aligned}$$

We rephrase (4.8), the upper bound in (4.10) for $r = 2$ and (4.11) in terms of Hecke eigenvalues: for a Hecke eigenform F and for $r \in \{1, 2, 4\}$ we have

$$(4.12) \quad \lambda(p^r, F)^2 = \sum_{0 \leq b \leq s \leq r} c_{r,b,s}(p) \lambda_{0,b}^{(2s)}(p, F), \quad c_{r,b,s}(p) \ll p^{3r-2s}.$$

¹which received rather negative reviews in mathscinet and Zentralblatt due to a somewhat sub-optimal presentation, but nevertheless the involved computations based on (4.2) are very useful

The existence of such a decomposition is obvious (recall (4.6)), the important point is the bound on the coefficients which is best-possible for $b = s$ (for smaller b better bounds are available, but we shall later need a uniform bound).

It is tempting to perform all of these computations in the algebra of Weyl group invariant polynomials using the Satake isomorphism. Unfortunately, the computations are by no means easier, since it is very hard to compute explicitly the image of a given double coset. As it may be useful in other situations, we have collected some explicit formulae in the appendix.

5. JOINT EIGENFUNCTIONS, REPRESENTATIONS AND SPECTRAL PARAMETERS

sec:SMF

A smooth function F on $\Gamma \backslash \mathcal{H}$ that is a joint eigenfunction of the algebra $\mathcal{D}(\mathcal{H})$ of G -invariant differential operators on \mathcal{H} and of moderate growth is commonly called a (Siegel) Maaß wave form for Γ . It is not necessarily square integrable, e.g. the (non-holomorphic) Eisenstein series constitute examples of not square integrable Maaß wave forms. However, if F additionally satisfies the regularity property

$$\int_{(\Gamma \cap N_j) \backslash N_j} F(n.Z) d_j n = 0$$

for all $Z \in \mathcal{H}$ and the unipotent radicals N_j (with Haar measure $d_j n$) of the proper parabolic subgroups P_j of G , then F is called cuspidal or a Siegel Maaß cusp form and it is in particular square integrable.

In order to apply the trace formula, we need a spectral decomposition of $L^2(\Gamma \backslash \mathcal{H})$ respecting the Hecke algebra. This is best done adelically, and we refer to [AS01] for a corresponding dictionary. Let \mathbb{A} be the adele ring of \mathbb{Q} . It follows from Langlands' monumental theory of Eisenstein series (see in particular [Lan76]), nicely summarized in [Art79], that the space $L^2(\mathrm{GSp}_4(\mathbb{Q}) \backslash \mathrm{GSp}_4(\mathbb{A}))$ has a $\mathrm{GSp}(\mathbb{A})$ -equivariant decomposition into a direct sum, parametrized by (classes of) parabolic subgroups and irreducible cuspidal automorphic representations of their Levi subgroups, of direct integrals. Each irreducible representation occurring in this decomposition is factorizable into local components, and for almost all places v , it contains a unique (up to scalars) K_v -fixed vector [Fla79]. In our situation, all representations are unramified at all finite places and hence generated by a Hecke eigenform.

We re-state in classical language that there exists a spectral decomposition

$$L^2(\Gamma \backslash \mathcal{H}) = L_{\mathrm{pp}}^2(\Gamma \backslash \mathcal{H}) \oplus L_{\mathrm{ac}}^2(\Gamma \backslash \mathcal{H})$$

where $L_{\mathrm{pp}}^2(\Gamma \backslash \mathcal{H})$ and $L_{\mathrm{ac}}^2(\Gamma \backslash \mathcal{H})$ are the subspaces corresponding to the pure point spectrum and the absolutely continuous spectrum, respectively. We write this decomposition as

spec1

$$(5.1) \quad L^2(\Gamma \backslash \mathcal{H}) = \int V_{\varpi} d\varpi$$

where each V_{ϖ} is a one-dimensional space generated by a (not necessarily square integrable) joint eigenfunction, i.e. an eigenfunction of the Hecke algebra \mathcal{H} and of the algebra $\mathcal{D}(\mathcal{H})$, in the sense that each function in the L^2 -space on the left hand side decomposes into a convergent sum and integral of functions from each subspace V_{ϖ} , and a corresponding Plancherel formula holds.

The space $L_{\mathrm{pp}}^2(\Gamma \backslash \mathcal{H})$ is spanned by the square integrable Siegel Maaß wave forms for Γ and decomposes further into

$$L_{\mathrm{pp}}^2(\Gamma \backslash \mathcal{H}) = L_{\mathrm{cusp}}^2(\Gamma \backslash \mathcal{H}) \oplus L_{\mathrm{res}}^2(\Gamma \backslash \mathcal{H}),$$

where $L_{\mathrm{cusp}}^2(\Gamma \backslash \mathcal{H})$ is spanned by the Siegel Maaß cusp forms, and $L_{\mathrm{res}}^2(\Gamma \backslash \mathcal{H})$ is spanned by square integrable iterated residues of the Eisenstein series (see [Kim95] for the determination of the residual spectrum).

We emphasize that the Hecke eigenvalues $\lambda(p, F)$ and $\lambda(p^2, F)$, and hence all $\lambda_{a,b}^{(r)}(p, F)$, of a joint eigenfunction $F \in V_{\varpi}$ are *real*. For L^2 -functions this follows trivially from the self-adjointness

of the Hecke operators with respect to the inner product (4.1). In general this is a consequence of the fact that the corresponding local representation ϖ_p is unitary, unramified and with trivial central character. If α_p, β_p denote the local Satake parameters, then $\lambda(p, F) = p^{3/2}(x + y)$ and $\lambda(p^2, F) = p^3(x^2 + xy + y^2)$ with $x = \alpha_p + \alpha_p^{-1}$, $y = \beta_p + \beta_p^{-1}$, and an inspection of [RoSc07, Table A.2] or [PS09, Proposition 3.1] shows that these numbers are real in all cases.

The joint eigenfunctions $F \in V_\varpi$ arise as vectors in an induced representation from the parabolic subgroup NA of the extension of a character

$$\chi: A \rightarrow \mathbb{C}^\times, \quad \text{diag}(e^{t_1}, e^{t_2}, e^{-t_1}, e^{-t_2}) \mapsto e^{it_1\mu_1 + it_2\mu_2}.$$

In this way we can identify the collection of functions in the various V_ϖ with a subset of $\mathfrak{a}_\mathbb{C}^*/W$, where we associate to each joint eigenfunction F the linear form (the spectral parameter) $\mu = \mu_1 e_1 + \mu_2 e_2 \in \mathfrak{a}_\mathbb{C}^*/W$ (or simply $\mathfrak{a}_\mathbb{C}^*$) that contains the (archimedean) Langlands parameters. By (3.4), the Laplace eigenvalue of F is then given by

$$\lambda_F = \frac{5 + \mu_1^2 + \mu_2^2}{12}. \quad (5.2)$$

A Weyl law of the form

$$\dim \text{span}\{F \in L_{\text{cusp}}^2(\Gamma \backslash \mathcal{H}) \mid \lambda_F \leq T\} \sim \text{const.} \cdot T^3$$

is known [LV07], rendering the statement of Theorem 1 to be non-void. We remark that in our normalization (see also (2.2) below) the tempered spectrum has real parameters $\mu \in \mathfrak{a}^*$. By [Nzo83], the spectral parameters (resp. representatives of their Weyl group orbits) of all irreducible unitary representations occurring in (5.1) are contained in

$$\Lambda := \mathfrak{a}^* \cup \{\mu = (\mu_1, \mu_2) \in \mathfrak{a}_\mathbb{C}^* : \|\text{Im } \mu\| \leq \sqrt{5/2} \text{ and } (\mu_1 = -\overline{\mu_2} \text{ or } \mu_1 \in \mathbb{R} \cup i\mathbb{R}, \mu_2 \in i\mathbb{R})\}. \quad (5.3)$$

(Unlike in the case of the group GL_2 , the set of exceptional parameters μ for a given lattice is not known to be finite.)

6. PRE-TRACE FORMULA AND AMPLIFICATION

The first step of Selberg's celebrated trace formula [Sel56] is the spectral expansion of an automorphic kernel, resulting in the following pre-trace formula. We recall the spectral decomposition (5.1) and denote by F_ϖ the normalized generator of V_ϖ and by $\mu_\varpi \in \Lambda \subseteq \mathfrak{a}_\mathbb{C}^*$ its spectral parameter. For any test function $f \in C_c^\infty(G//K)$ we have

$$\int \tilde{f}(\mu_\varpi) F_\varpi(x) \overline{F_\varpi(y)} d\varpi = \sum_{\gamma \in \Gamma} f(x^{-1}\gamma y) \quad \text{for } x, y \in G, \quad (6.1)$$

where \tilde{f} is the spherical transform defined in (2.3). The functions F_ϖ are understood here as right- K -invariant functions on $\Gamma \backslash G$. Let $F_0 \in L^2(\Gamma \backslash \mathcal{H})$ be the L^2 -normalized joint eigenfunction with spectral parameter $\mu_0 \in \Lambda$ whose supremum norm we want to bound.

6.1. Choice of test function. In order to prove Theorem 1, we choose a specific test function $f \in C_c^\infty(G//K)$ to be used in (6.1) adapted to the spectral parameter μ_0 . The choice of this test function is done indirectly by choosing its spherical transform $\mathcal{H}_W(\mathfrak{a}_\mathbb{C}^*)$ with sufficiently small support about the identity $o \in G//K$ and applying an inverse transformation. We prove upper bounds on f using Proposition 3 in connection with Theorem 2. Recall that without loss of generality we may assume $\|\mu_0\|$ to be sufficiently large. The test function and its spherical transform will only depend on the real part

$$\mu := \text{Re } \mu_0 \in \mathfrak{a}^* \quad (6.2)$$

of the spectral parameter of F_0 .

Let $\psi \in \mathcal{H}_W^1(\mathfrak{a}_{\mathbb{C}}^*)$ be a fixed Paley–Wiener function² such that

- (a) ψ is nonnegative on \mathfrak{a}^* , so in particular $\psi(\bar{\lambda}) = \overline{\psi(\lambda)}$,
- (b) $\operatorname{Re} \psi \geq 1$ in a ball of radius $\sqrt{5/2}$ about $0 \in \mathfrak{a}_{\mathbb{C}}^*$.

For $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$ we define

$$\text{defftilde} \quad (6.3) \quad \tilde{f}_{\mu}(\lambda) := \left(\sum_{w \in W} \psi(\mu - w.\lambda) \right)^2.$$

Clearly, $\tilde{f}_{\mu} \in \mathcal{H}_W^2(\mathfrak{a}_{\mathbb{C}}^*)$. Further, it is easy to see that

$$\text{pos} \quad (6.4) \quad \tilde{f}_{\mu}(\lambda) \geq 0 \text{ for all } \lambda \in \Lambda$$

and

$$\text{big} \quad (6.5) \quad \tilde{f}_{\mu}(\lambda) \geq 1 \text{ for all } \lambda \in \Lambda \text{ whenever } \operatorname{Re} \lambda = \mu.$$

The latter is obvious for $\lambda \in \mathfrak{a}^*$. We recall the classification of possible exceptional parameters in (5.3). For an exceptional spectral parameter of the form $\lambda = (x + iy)e_1 + (-x + iy)e_2$ and $\mu = xe_1 - xe_2$ with $x \in \mathbb{R}$ sufficiently large and $y \in [-\sqrt{5/2}, \sqrt{5/2}]$ we have

$$\begin{aligned} \tilde{f}_{\mu}(\lambda) &= 4 \left(\operatorname{Re} \psi(i(y, y)) + \operatorname{Re} \psi((2x, 0) + i(-y, y)) \right. \\ &\quad \left. + \operatorname{Re} \psi((0, -2x) + i(-y, y)) + \operatorname{Re} \psi((2x, -2x) + i(y, y)) \right)^2 \\ &\geq 3 |\operatorname{Re} \psi(i(y, y))|^2 \geq 1 \end{aligned}$$

by the rapid decay of ψ for large x and $\psi(\bar{\lambda}) = \overline{\psi(\lambda)}$. For an exceptional spectral parameter of the form $\lambda = iye_1 + xe_2$ and $\mu = xe_2$ we have by the same argument

$$\begin{aligned} \tilde{f}_{\mu}(\lambda) &= 4 \left(\operatorname{Re} \psi(i(0, y)) + \operatorname{Re} \psi((x, x) + i(y, 0)) \right. \\ &\quad \left. + \operatorname{Re} \psi((2x, 0) + i(0, y)) + \operatorname{Re} \psi((x, -x) + i(y, 0)) \right)^2 \\ &\geq 3 |\operatorname{Re} \psi(i(0, y))|^2 \geq 1. \end{aligned}$$

Exceptional spectral parameters with λ_1, λ_2 both purely imaginary are bounded and hence do not meet the condition $\operatorname{Re} \lambda = \mu$.

Finally, $\tilde{f}_{\mu} \in \mathcal{H}_W(\mathfrak{a}_{\mathbb{C}}^*)$ immediately implies

$$\text{decay} \quad (6.6) \quad \tilde{f}_{\mu}(\lambda) \ll_A \max_{w \in W} (1 + \|\mu - w.\lambda\|)^{-A}$$

for $\lambda \in \mathfrak{a}^*$.

By Proposition 3 the inverse spherical transform f_{μ} of \tilde{f}_{μ} has support in $\overline{B_2^A(o)}$, independent of μ . Combining Theorem 2, Proposition 3, (3.2) and (6.6) and recalling the notation (2.1) for the Cartan projection, we conclude

$$\text{mainbound} \quad (6.7) \quad f_{\mu}(g) \ll \|\mu\|^4 (1 + \|\mu\| \cdot \|C(g)\|)^{-1/2}.$$

²by choosing $\psi \in \mathcal{H}_W^{C_0/2}(\mathfrak{a}_{\mathbb{C}}^*)$ where C_0 is the global constant appearing in the proof of Theorem 2, we could have made the argument independent of [DKV83].

6.2. Construction of the amplifier. With the choice (6.3) we return to (6.1) and construct a suitable amplifier. Given a double coset $\Gamma M \Gamma = \bigcup_i \Gamma M_i$ with $M, M_i \in \mathrm{GSp}_4^+(\mathbb{Z})$, we apply (6.1) for the elements $(x, y) = (g, (\det M_i)^{-1/4} M_i g) \in G \times G$ obtaining

$$\int \lambda(M, F_\varpi) \tilde{f}_\mu(\mu_\varpi) |F_\varpi(g)|^2 d\varpi = \sum_{\gamma \in \Gamma M \Gamma} f_\mu(g^{-1} \tilde{\gamma} g)$$

where $\tilde{\gamma} := (\det \gamma)^{-1/4} \gamma$ and $\lambda(M, F_\varpi)$ is the eigenvalue of F_ϖ with respect to $T_{\Gamma M \Gamma}$.

Let $L \geq 5$ be a parameter and let \mathcal{P} be the set of primes in $(L, 2L]$. Define $x(n) := \mathrm{sgn}(\lambda(n, F_0))$ for $n \in \mathbb{N}$ and let

$$A_\varpi := \left(\sum_{l \in \mathcal{P}} x(l) \lambda(l, F_\varpi) \right)^2 + \left(\sum_{l \in \mathcal{P}} x(l^2) l^{-3/2} \lambda(l^2, F_\varpi) \right)^2 + \left(\sum_{l \in \mathcal{P}} x(l^4) l^{-9/2} \lambda(l^4, F_\varpi) \right)^2 \geq 0.$$

By (4.7) and the prime number theorem we have

$$\begin{aligned} A_0 &= \left(\sum_{l \in \mathcal{P}} |\lambda(l, F_0)| \right)^2 + \left(\sum_{l \in \mathcal{P}} l^{-3/2} |\lambda(l^2, F_0)| \right)^2 + \left(\sum_{l \in \mathcal{P}} l^{-9/2} |\lambda(l^4, F_0)| \right)^2 \\ &\geq \frac{1}{3} \left(\sum_{l \in \mathcal{P}} |\lambda(l, F_0)| + l^{-3/2} |\lambda(l^2, F_0)| + l^{-9/2} |\lambda(l^4, F_0)| \right)^2 \gg L^5 (\log L)^{-2}. \end{aligned} \quad (6.8)$$

Combining (4.4) and (4.12), we can expand A_ϖ :

$$\begin{aligned} A_\varpi &= \sum_{l_1 \neq l_2} \left(x(l_1 l_2) \lambda(l_1 l_2, F_\varpi) + \frac{x(l_1^2 l_2^2)}{(l_1 l_2)^{3/2}} \lambda(l_1^2 l_2^2, F_\varpi) + \frac{x(l_1^4 l_2^4)}{(l_1 l_2)^{9/2}} \lambda(l_1^4 l_2^4, F_\varpi) \right) \\ &\quad + \sum_{0 \leq b \leq s \leq 4} \sum_l \xi_{b,s}(l) \lambda_{0,b}^{(2s)}(l, F_\varpi) \end{aligned} \quad (6.9)$$

where $\xi_{b,s}(l) \ll l^{3-2s}$. By (6.4), (6.5), (6.8) and (6.9) we conclude that

$$\begin{aligned} \frac{L^5}{(\log L)^2} |F_0(g)|^2 &\ll \int A_\varpi \tilde{f}_\mu(\mu_\varpi) |F_\varpi(g)|^2 d\varpi \\ &\ll \sum_{r \in \{1,2,4\}} \sum_{l_1 \neq l_2 \in \mathcal{P}(L)} (l_1 l_2)^{\frac{3}{2}(1-r)} \sum_{\gamma \in S(l_1^r l_2^r)} |f_\mu(g^{-1} \tilde{\gamma} g)| \\ &\quad + \sum_{0 \leq r \leq 4} \sum_{l \in \mathcal{P}(L)} l^{3-2r} \sum_{\gamma \in S(l^{2r})} |f_\mu(g^{-1} \tilde{\gamma} g)| \end{aligned} \quad (6.10)$$

with $S(m)$ as in (4.3).

6.3. Proof of Theorem 1. In order to prove Theorem 1, we estimate the right hand side of (6.10). To this end, we fix some small $0 < \delta_0 < 1$. We estimate the number of matrices γ with $\|C(g^{-1} \tilde{\gamma} g)\| \geq \delta_0$ trivially and obtain a saving from the decay of f_μ as specified in (6.7). On the other hand, we will prove a strong bound for the number of matrices γ with $\|C(g^{-1} \tilde{\gamma} g)\| \leq \delta_0$, and for those we estimate f_μ trivially. With this in mind, we define

$$\mathcal{S}(g)_\delta[m] := \{ \gamma \in S(m) \mid \|\tilde{\gamma} - gKg^{-1}\| \leq \delta \} \quad (6.11)$$

for $m \in \mathbb{N}$ and $\delta > 0$. Since g varies in a fixed compact set Ω , we make the observation that $\|C(g^{-1} \tilde{\gamma} g)\| \leq \delta$ for some $\delta \ll 1$ and $\det \gamma = m^2$ imply $\tilde{\gamma} \in gKg^{-1} + O_\Omega(\delta)$ and hence $\gamma \in \mathcal{S}(g)_{c\delta}[m]$ for some constant $c > 0$ depending on Ω .

The next section is devoted to a bound for the cardinality of $\mathcal{S}(g)_\delta[m]$ uniformly in δ and m . Our principal result in this direction is

Proposition 5. *There exist $\eta, B > 0$ such that for all $\delta \ll 1$, $\varepsilon > 0$, $m \in \mathbb{N}$ and $g \in G$ we have*

$$\#\mathcal{S}(g)_\delta[m] \ll_{g,\varepsilon} m^{1+\varepsilon} (1 + \delta^\eta m^B).$$

The implied constant does not depend on m and δ .

Taking this for granted (a proof follows in Section 7 below), it is now a simple matter to prove Theorem 1.

Proof of Theorem 1. We insert (6.7) and Proposition 5 into (6.10). The constant B used below is global and possibly larger than the one in Proposition 5, the variables δ, L will be fixed below. As mentioned before, we can assume that the spectral parameter μ_0 (and hence its real part μ) of our form F_0 is sufficiently large.

The contribution of the matrices γ with $\|C(g^{-1}\tilde{\gamma}g)\| \geq \delta$ to the right hand side of (6.10) is

$$\ll_{\Omega} \|\mu\|^{7/2} \delta^{-1/2} L^B,$$

and the contribution of the matrices γ with $\|C(g^{-1}\tilde{\gamma}g)\| \leq \delta$ to the right hand side of (6.10) is

$$\ll_{\varepsilon, \Omega} \|\mu\|^4 L^{4+\varepsilon} (1 + \delta^\eta L^B).$$

Upon choosing

$$\delta^{1/2+\eta} := \|\mu\|^{-1/2} \quad \text{and} \quad L := \left\lceil \delta^{-\frac{\eta}{B}} \right\rceil$$

we obtain from (6.10) that

$$|F_0(g)|^2 \ll_{\Omega, \varepsilon} \|\mu\|^4 L^{-1+\varepsilon} (\log L)^2 \ll \|\mu\|^4 L^{-\frac{3}{4}}.$$

Thus,

$$F_0(g) \ll_{\varepsilon, \Omega} \|\mu\|^{2 - \frac{3\eta}{4B(1+2\eta)} + \varepsilon}.$$

Theorem 1 now follows from (5.2). □

sec7

7. DIOPHANTINE ANALYSIS

It remains to prove Proposition 5 which is the purpose of this section. Let

$$k = \begin{pmatrix} B & C \\ -C & B \end{pmatrix} \in \sqrt{m}K.$$

This matrix is symplectic and orthogonal, i.e.

$$(7.1) \quad k^\top k = mI_4 \quad \text{and} \quad k^\top Jk = mJ.$$

Let

$$Q := (gg^\top)^{-1},$$

which is a positive definite matrix depending only on $g \in G$. To simplify notation we set

$$Q = (q_{ij}) = (q_1 \ q_2 \ q_3 \ q_4),$$

where $q_i = (q_{1i}, q_{2i}, q_{3i}, q_{4i})^\top$ are column vectors. Further we set

$$Q_1 := (q_1 \ 0 \ q_3 \ 0) \quad \text{and} \quad Q_2 := (0 \ q_2 \ 0 \ q_4),$$

hence $Q = Q_1 + Q_2$.

propstab1

Proposition 6. *Let $\delta \ll 1$. For any $g \in G$, the set $\mathcal{S}(g)_\delta[m]$ defined in (6.11) consists of matrices*

$$\gamma = \begin{pmatrix} r_1 & s_1 & * & * \\ r_2 & s_2 & * & * \\ r_3 & s_3 & * & * \\ r_4 & s_4 & * & * \end{pmatrix}$$

with the following properties:

- (a) All entries of the matrix are $\ll_g m^{1/2}$.
- (b) The vectors $r = (r_1, r_2, r_3, r_4)^\top$ and $s = (s_1, s_2, s_3, s_4)^\top$ determine each of the remaining entries of the matrix up to $O_g(\delta m^{1/2})$.
- (c) We have

$$(7.2) \quad r^\top Qr = mq_{11} + O_g(\delta m), \quad s^\top Qs = mq_{22} + O_g(\delta m).$$

propstab1a

propstab1b

propstab1c

mat1a

propstab1d (d) If $r^\top Q_1 r \asymp m$, then

$$\begin{aligned} s_1 &= \frac{s_2 \cdot r^\top A_{11} r + s_4 \cdot r^\top A_{12} r + m r_1 q_{12}}{r^\top Q_1 r} + O_g(\delta m^{1/2}), \\ s_3 &= \frac{s_2 \cdot r^\top A_{21} r + s_4 \cdot r^\top A_{22} r + m r_3 q_{12}}{r^\top Q_1 r} + O_g(\delta m^{1/2}) \end{aligned} \quad (7.3)$$

with $A_{11} = (-q_2 \ 0 \ 0 \ -q_3)$, $A_{12} = (-q_4 \ q_3 \ 0 \ 0)$, $A_{21} = (0 \ 0 \ -q_2 \ q_1)$ and $A_{22} = (0 \ -q_1 \ -q_4 \ 0)$.
If $r^\top Q_2 r \asymp m$, then

$$\begin{aligned} s_2 &= \frac{s_1 \cdot r^\top B_{11} r + s_3 \cdot r^\top B_{12} r + m r_2 q_{12}}{r^\top Q_2 r} + O_g(\delta m^{1/2}), \\ s_4 &= \frac{s_1 \cdot r^\top B_{21} r + s_3 \cdot r^\top B_{22} r + m r_4 q_{12}}{r^\top Q_2 r} + O_g(\delta m^{1/2}) \end{aligned} \quad (7.4)$$

with $B_{11} = (0 \ -q_1 \ -q_4 \ 0)$, $B_{12} = (q_4 \ -q_3 \ 0 \ 0)$, $B_{21} = (0 \ 0 \ q_2 \ -q_1)$ and $B_{22} = (-q_2 \ 0 \ 0 \ -q_3)$.

Proof. Part (a) is obvious from the definition (6.11).

Part (b) follows from

$$gkg^{-1} = g \begin{pmatrix} B & C \\ -C & B \end{pmatrix} g^{-1} = \begin{pmatrix} VBV^{-1} - XV^{-1}CV^{-1} & * \\ -V^{-1}CV^{-1} & * \end{pmatrix}$$

where we used the notation (3.3). Hence for fixed X, V , the first two columns of a matrix $\gamma \in m^{1/2}gKg^{-1} + O(\delta m^{1/2})$ determine B and C up to $O(\delta m^{1/2})$, and hence the remaining two columns up to $O(\delta m^{1/2})$.

By (7.1) and part (a) we have $\gamma^\top Q \gamma = mQ + O_g(\delta m)$ for $\gamma \in \mathcal{S}(g)_\delta[m]$ which implies part (c).

The same argument shows $\gamma^\top J \gamma = mJ + O_g(\delta m)$ for $\gamma \in \mathcal{S}(g)_\delta[m]$, hence

$$(7.5) \quad r^\top J s = O_g(\delta m) \quad \text{and} \quad r^\top Q s = m q_{12} + O_g(\delta m).$$

If $r^\top Q_1 r \asymp m$ (in particular $\neq 0$), we solve the linear system (7.5) for s_1 and s_3 , obtaining (7.3). If $r^\top Q_2 r \asymp m$, the same argument gives (7.4). This completes the proof of part (d). \square

We see that for sufficiently small δ , the matrices in question are essentially characterized by the 8 variables r_j, s_j that satisfy the four quadratic equations (7.2) and (7.5). Hence geometrically we need to study the lattice points in a δ -neighbourhood of the intersection of four quadrics in 8 variables. Roughly speaking, we will choose r_1, r_2 freely, then r_3, r_4 are essentially determined by the first equality in (7.2) and the theory of binary quadratic forms. Once r is fixed, we substitute (7.3) or (7.4) into the second equation of (7.2), obtaining again a binary problem that essentially fixes s . Hence the cardinality of the integral matrices in $\mathcal{S}(g)_\delta[m]$ is $O(m^{1+\varepsilon})$ for sufficiently small δ . At least for very small δ and odd positive integers m , the bound of Proposition 5 is essentially best possible, since $\mathcal{S}(\text{id})_\delta[m]$ contains, for every $\delta > 0$, all the $\asymp m$ matrices of the form

$$\begin{pmatrix} a_1 & a_3 & a_2 & a_4 \\ -a_3 & a_1 & a_4 & -a_2 \\ -a_2 & -a_4 & a_1 & a_3 \\ -a_4 & a_2 & -a_3 & a_1 \end{pmatrix}, \quad a_1^2 + a_2^2 + a_3^2 + a_4^2 = m.$$

In order to make these arguments rigorous in the following we start with a multi-dimensional version of Dirichlet's approximation theorem [HW54, Theorem 200].

lem1 **Lemma 7.** *Let ξ_1, \dots, ξ_n be real numbers, $T > 1$. Then there exist integers p_1, \dots, p_n and a positive integer $q \leq T$ such that $|\xi_j - p_j/q| \leq (qT^{1/n})^{-1}$ for all $1 \leq j \leq n$.*

For a polynomial P we denote by $H(P)$ the largest coefficient in absolute value.
The next lemma is standard.

lem2a

Lemma 8. (a) Let $P(x, y) \in \mathbb{Z}[x, y]$ be a quadratic polynomial whose quadratic homogeneous part is positive definite. Then the number of integral solutions to $P(x, y) = 0$ is $O_\varepsilon(H(P)^\varepsilon)$ for all $\varepsilon > 0$.

lem2b

(b) There exists a constant $C > 0$ such that for each $\delta, D > 0$ and each quadratic polynomial $P(x, y) \in \mathbb{R}[x, y]$ whose quadratic homogeneous part is positive definite with discriminant $|\Delta| \geq D$, the bound $|P(x, y)| \leq \delta$ implies $\max(|x|, |y|) \ll_D (\delta + 1 + H(P))^C$.

Proof. We write $P(x, y) = ax^2 + bxy + cy^2 + dx + ey + f$ and $\Delta = b^2 - 4ac < 0$ (so in particular $a \neq 0$). We write $\xi = (be - 2cd)/\Delta$ and $\eta = (bd - 2ae)/\Delta$. One checks that

$$(7.6) \quad P(x, y) = \frac{(2a(x + \xi) + b(y + \eta))^2 - \Delta(y + \eta)^2}{4a} + P(-\xi, -\eta).$$

Hence $P(x, y) = 0$ implies $X^2 - \Delta Y^2 = -4a\Delta^2 P(-\xi, -\eta)$ for certain integers X, Y . The number of solutions in X, Y is at most the number of ideals in $\mathbb{Q}(\sqrt{\Delta})$ of norm $-4a\Delta^2 P(-\xi, -\eta)$ which is bounded by 6 times the number of divisors of $|4a\Delta^2 P(-\xi, -\eta)|$. The well-known growth bound $d(n) = O_\varepsilon(n^\varepsilon)$ for the divisor function now implies part (a). Part (b) follows from straightforward estimates: first

$$\delta \geq |P(x, y)| \geq \frac{|\Delta||y + \eta|^2}{4|a|} - |P(-\xi, -\eta)|$$

implies $|y| \ll_D (\delta + 1 + H(P))^C$. Using this bound in (7.6) yields then the claimed bound for $|x|$. \square

cor3

Corollary 9. There exists a constant $A > 0$ such that for each $\varepsilon, \delta, D > 0$ and each quadratic polynomial $P(x, y) \in \mathbb{R}[x, y]$ whose quadratic homogeneous part is positive definite with discriminant $|\Delta| \geq D$ we have

$$\#\{(x, y) \in \mathbb{Z}^2 \mid |P(x, y)| < \delta\} \ll_{D, \varepsilon} Z^\varepsilon + \delta^{1/7} Z^A$$

where $Z = \delta + 1 + H(P)$.

Proof. Let $(x, y) \in \mathbb{Z}^2$ with $|P(x, y)| < \delta$. By Lemma 8(b) we can assume that $\max(|x|, |y|) \ll_D Z^C$. Let $T > 1$ be a parameter to be chosen later. We approximate the six coefficients of P by rational numbers with common denominator $q \leq T$. Using Lemma 7 and multiplying by q , we obtain

$$|\tilde{P}(x, y)| \leq c_D(\delta T + Z^{2C} T^{-1/6}) =: R$$

for some integral polynomial $\tilde{P} \in \mathbb{Z}[x, y]$ of height $H(\tilde{P}) \ll_D TH(P)$ and a constant $c_D > 0$ only depending on D . For each integer $r \leq R$ we bound the number of solutions to $\tilde{P}(x, y) - r = 0$ by Lemma 8(a) getting a total of $\ll_{\varepsilon, D} (1 + R)(R + TH(P))^\varepsilon$ solutions at most. We choose

$$T = 1 + \min(Z^{12C/7} \delta^{-6/7}, Z^{12C}),$$

so that $1 + R \ll_D 1 + \delta + \delta^{1/7} Z^{12C/7}$ and $R + TH(P) = Z^{O(1)}$. This completes the proof. \square

We are now prepared for the

Proof of Proposition 5. We choose $r_1, r_2 \ll_g m^{1/2}$, and substitute the values into the first equation of (7.2). This provides us with the quadratic polynomial

$$\begin{aligned} P(r_3, r_4) &= q_{44}r_4^2 + 2q_{34}r_3r_4 + q_{33}r_3^2 + 2(q_{13}r_1 + q_{23}r_2)r_3 + 2(q_{14}r_1 + q_{24}r_2)r_4 \\ &\quad + q_{11}r_1^2 + 2q_{12}r_1r_2 + q_{22}r_2^2 - mq_{11}, \end{aligned}$$

whose quadratic homogeneous part is positive definite and $H(P) \ll_g m$. By (7.2), $|P(r_3, r_4)| \ll_g \delta m$. Hence Corollary 9 shows that we have $\ll_{g, \varepsilon} m^\varepsilon + (\delta m)^{1/7} m^A$ choices for r_3, r_4 . Without loss of generality let us assume that $r^\top Q_1 r \asymp m$. We substitute (7.3) into the second equation of (7.2) and get a binary quadratic form $\tilde{P}_r(s_2, s_4)$ whose coefficients, and also its discriminant, depend on r (and g). Since Q is positive definite, so is \tilde{P}_r . Moreover, its shortest vector is trivially bounded below by the shortest vector of Q which is bounded below by a constant depending only on g . Minkowski's lower bound for the discriminant of a quadratic form by its successive minima [Cas78, Theorem 2.2]

now shows that the discriminant of the quadratic homogeneous part of \tilde{P}_r is bounded away from 0 uniformly in r . Clearly, $H(\tilde{P}_r) \ll_g m$. Then Corollary 9 restricts the number of choices for s_2, s_4 to $\ll_{g,\varepsilon} m^\varepsilon + (\delta m)^{1/7} m^A$. Now s_1, s_3 and the remaining 8 entries are determined up to $O_g(\delta m^{1/2})$. This gives a total count for $\#\mathcal{S}(g)_\delta[m]$ of $\ll_{g,\varepsilon} m^{1+\varepsilon} (1 + (\delta m)^{2/7} m^{2A}) (1 + \delta m^{1/2})^{10}$. \square

8. APPENDIX: IMAGES OF HECKE OPERATORS UNDER THE SATAKE MAP

Given a double coset $\Gamma \text{diag}(p^a, p^b, p^{r-b}, p^{r-a}) \Gamma$, there exists a decomposition into left cosets

$$\Gamma M \Gamma = \bigcup_j \Gamma M_j, \quad M_j = \begin{pmatrix} A_j & * \\ p^r A_j^{-\top} & \end{pmatrix}, \quad A_j = \begin{pmatrix} p^\alpha & * \\ & p^\beta \end{pmatrix}.$$

The image of $T_{\Gamma M \Gamma}$ under the Satake map is the polynomial

$$x_0^r \sum_j \left(\frac{x_1}{p} \right)^\alpha \left(\frac{x_2}{p^2} \right)^\beta \in \mathbb{C}[x_0, x_1, x_2].$$

This map is an algebra isomorphism between the p -part of the integral Hecke algebra and polynomials that are symmetric in x_1, x_2 and invariant under the automorphisms

$$(x_0, x_1, x_2) \mapsto (x_0 x_1, 1/x_1, x_2) \quad \text{and} \quad (x_0, x_1, x_2) \mapsto (x_0 x_2, x_1, 1/x_2).$$

Table 1 is compiled using the results of [Kod67, p. 120] and comparing coefficients. Following [RySh08], it is most efficient to use symmetrized polynomials. For $\mathbf{a} = (a_1, a_2)$ with $0 \leq a_1 \leq a_2 \leq r/2$ we define the Weyl orbit as

$$W_r(a_1, a_2) := \{(a_1, a_2), (a_2, a_1), (r - a_1, a_2), (a_2, r - a_1), \\ (a_1, r - a_2), (r - a_2, a_1), (r - a_1, r - a_2), (r - a_2, r - a_1)\}$$

and

$$\mathbf{x}_r^{\mathbf{a}} = \sum_{(b_1, b_2) \in W_r(a_1, a_2)} x_1^{b_1} x_2^{b_2}.$$

The entries for the Hecke operators of order 3, 4, 5, 6 do not seem to be in the literature in explicit form and extend the matrix in [RySh08, p. 238].

$T_{0,0}^{(1)}$	$x_0 \mathbf{x}_1^{(0,0)}$
$T_{0,0}^{(2)}$	$x_0^2 \left(\mathbf{x}_2^{(0,0)} + \frac{p-1}{p} \mathbf{x}_2^{(0,1)} + \frac{2(p-1)}{p} \mathbf{x}_2^{(1,1)} \right)$
$T_{0,1}^{(2)}$	$x_0^2 \left(\frac{1}{p} \mathbf{x}_2^{(0,1)} + \frac{p^2-1}{p^3} \mathbf{x}_2^{(1,1)} \right)$
$T_{0,0}^{(3)}$	$x_0^3 \left(\mathbf{x}_3^{(0,0)} + \frac{p-1}{p} \mathbf{x}_3^{(0,1)} + \frac{(p-1)(2p-1)}{p^2} \mathbf{x}_3^{(1,1)} \right)$
$T_{0,1}^{(3)}$	$x_0^3 \left(\frac{1}{p} \mathbf{x}_3^{(0,1)} + \frac{(p-1)(2p+1)}{p^3} \mathbf{x}_3^{(1,1)} \right)$
$T_{0,0}^{(4)}$	$x_0^4 \left(\mathbf{x}_4^{(0,0)} + \frac{p-1}{p} \mathbf{x}_4^{(0,1)} + \frac{p-1}{p} \mathbf{x}_4^{(0,2)} + \frac{(p-1)(2p-1)}{p^2} \mathbf{x}_4^{(1,1)} + \frac{2(p-1)^2}{p^2} \mathbf{x}_4^{(1,2)} + \frac{(p-1)(3p^2-2p+1)}{p^3} \mathbf{x}_4^{(2,2)} \right)$
$T_{0,1}^{(4)}$	$x_0^4 \left(\frac{1}{p} \mathbf{x}_4^{(0,1)} + \frac{p-1}{p} \mathbf{x}_4^{(0,2)} + \frac{2(p-1)}{p^2} \mathbf{x}_4^{(1,1)} + \frac{3(p-1)}{p^2} \mathbf{x}_4^{(1,2)} + \frac{(p-1)^2(3p+1)}{p^4} \mathbf{x}_4^{(2,2)} \right)$
$T_{0,2}^{(4)}$	$x_0^4 \left(\frac{1}{p^2} \mathbf{x}_4^{(0,2)} + \frac{p-1}{p^3} \mathbf{x}_4^{(1,1)} + \frac{p-1}{p^3} \mathbf{x}_4^{(1,2)} + \frac{2(p-1)}{p^3} \mathbf{x}_4^{(2,2)} \right)$
$T_{0,0}^{(5)}$	$x_0^5 \left(\mathbf{x}_5^{(0,0)} + \frac{p-1}{p} \mathbf{x}_5^{(0,1)} + \frac{p-1}{p} \mathbf{x}_5^{(0,2)} + \frac{(p-1)(2p-1)}{p^2} \mathbf{x}_5^{(1,1)} + \frac{2(p-1)^2}{p^2} \mathbf{x}_5^{(1,2)} + \frac{(p-1)(3p^2-3p+1)}{p^3} \mathbf{x}_5^{(2,2)} \right)$
$T_{0,1}^{(5)}$	$x_0^5 \left(\frac{1}{p} \mathbf{x}_5^{(0,1)} + \frac{p-1}{p} \mathbf{x}_5^{(0,2)} + \frac{2(p-1)}{p^2} \mathbf{x}_5^{(1,1)} + \frac{(3p-1)(p-1)}{p^3} \mathbf{x}_5^{(1,2)} + \frac{(p-1)(4p-3)}{p^3} \mathbf{x}_5^{(2,2)} \right)$
$T_{0,2}^{(5)}$	$x_0^5 \left(\frac{1}{p^2} \mathbf{x}_5^{(0,2)} + \frac{p-1}{p^3} \mathbf{x}_5^{(1,1)} + \frac{2(p-1)}{p^3} \mathbf{x}_5^{(1,2)} + \frac{(p-1)(3p-1)}{p^3} \mathbf{x}_5^{(2,2)} \right)$
$T_{0,0}^{(6)}$	$x_0^6 \left(\mathbf{x}_6^{(0,0)} + \frac{p-1}{p} (\mathbf{x}_6^{(0,1)} + \mathbf{x}_6^{(0,2)} + \mathbf{x}_6^{(0,3)}) + \frac{(p-1)(2p-1)}{p^2} \mathbf{x}_6^{(1,1)} + \frac{2(p-1)^2}{p^2} (\mathbf{x}_6^{(1,2)} + \mathbf{x}_6^{(1,3)}) \right. \\ \left. + \frac{(p-1)(3p^2-3p+1)}{p^3} \mathbf{x}_6^{(2,2)} + \frac{(p-1)^2(3p-1)}{p^3} \mathbf{x}_6^{(2,3)} + \frac{2(p-1)(2p^2-2p+1)}{p^3} \mathbf{x}_6^{(3,3)} \right)$
$T_{0,1}^{(6)}$	$x_0^6 \left(\frac{1}{p} \mathbf{x}_6^{(0,1)} + \frac{p-1}{p^2} (\mathbf{x}_6^{(0,2)} + \mathbf{x}_6^{(0,3)}) + \frac{2(p-1)}{p^2} \mathbf{x}_6^{(1,1)} + \frac{(p-1)(3p-1)}{p^3} \mathbf{x}_6^{(1,2)} + \frac{(p-1)(3p-2)}{p^3} \mathbf{x}_6^{(1,3)} \right. \\ \left. + \frac{4(p-1)^2}{p^3} \mathbf{x}_6^{(2,2)} + \frac{(p-1)(5p^2-4p+1)}{p^4} \mathbf{x}_6^{(2,3)} + \frac{(p-1)^2(5p-1)}{p^4} \mathbf{x}_6^{(3,3)} \right)$

TABLE 1. Polynomials for Hecke operators under the Satake map

extension

REFERENCES

- [AZ95] A. Andrianov and V. Zhuravlev, *Modular forms and Hecke operators*, American Math. Society, 1995.
- [Art79] J. Arthur, *Eisenstein series and the trace formula*, Automorphic forms, representations and L-functions, Corvallis/Oregon 1977, Proc. Symp. Pure Math. 33, vol. 1, 253-274, 1979.
- [AS01] M. Asgari and R. Schmidt, *Siegel modular forms and representations*, Manuscr. Math. **104** (2001), 173–200.
- [BHM] V. Blomer, G. Harcos, and D. Milićević, *Eigenfunctions on arithmetic hyperbolic 3-manifolds*, arXiv:1401.5154
- [BM13] V. Blomer and P. Michel, *Hybrid bounds for automorphic forms on ellipsoids over number fields*, J. Inst. Math. Jussieu **12** (2013), no. 4, 727–758.
- [BvdGHZ08] J. Bruinier, G. van der Geer, G. Harder, and D. Zagier, *The 1-2-3 of modular forms. Lectures at a summer school in Nordfjordeid, Norway, June 2004*, Berlin: Springer, 2008.
- [BT] F. Brumley and N. Templier, *Large values of cusp forms on GL(n)*, preprint.
- [Cas78] J. Cassels, *Rational quadratic forms*, L.M.S. Monographs, No.13, 1978.
- [DKV83] J. Duistermaat, J. Kolk and V. Varadarajan, *Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups*, Compositio Math. **49** (1983), no. 3, 309–398.
- [Far08] J. Faraut, *Analysis on Lie groups. An introduction*, Cambridge University Press, 2008.
- [Fla79] D. Flath, *Decomposition of representations into tensor products*, Automorphic forms, representations and L-functions, Corvallis/Oregon 1977, Proc. Symp. Pure Math. 33, vol. 1, 179-183, 1979.
- [Fre83] E. Freitag, *Siegelsche Modulfunktionen*, Grundlehren der Mathematischen Wissenschaften, vol. 254, Springer-Verlag, Berlin, 1983.
- [Gan71] R. Gangolli, *On the Plancherel formula and the Paley-Wiener theorem for spherical functions on semisimple Lie groups*, Ann. Math. (2) **93** (1971), 150–165.
- [HT13] G. Harcos and N. Templier, *On the sup-norm of Maass cusp forms of large level. III*, Math. Ann. **356** (2013), no. 1, 209–216.
- [HW54] G. Hardy and E. Wright, *An introduction to the theory of numbers. 3rd ed.*, Oxford: At the Clarendon Press. xvi+419 pp., 1954.

- [Hel84] S. Helgason, *Groups and geometric analysis*, Integral geometry, invariant differential operators, and spherical functions, Pure and Applied Mathematics, vol. 113, Academic Press Inc., Orlando, FL, 1984.
- [Hel08] ———, *Geometric analysis on symmetric spaces*, 2nd ed., Providence, RI: American Mathematical Society, 2008.
- [HJ69] S. Helgason and K. Johnson, *The bounded spherical functions on symmetric spaces*, Adv. Math. **3** (1969), 586–593.
- [Hoe68] L. Hörmander, *The spectral function of an elliptic operator*, Acta Math. **121** (1968), 193–218.
- [HL94] J. Hoffstein and P. Lockhart, *Coefficients of Maass forms and the Siegel zero. Appendix: An effective zero-free region, by Dorian Goldfeld, Jeffrey Hoffstein and Daniel Lieman*, Ann. of Math. **140** (1994), 161–176.
- [Iwa97] H. Iwaniec, *Topics in classical automorphic forms*, Providence, RI: Amer. Math. Soc., 1997.
- [IS95] H. Iwaniec and P. Sarnak, *L^∞ norms of eigenfunctions of arithmetic surfaces*, Ann. Math. (2) **141** (1995), no. 2, 301–320.
- [Kim95] H. Kim, *The residual spectrum of Sp_4* , Compos. Math. **99** (1995), no. 2, 129–151.
- [Kna02] A. Knapp, *Lie groups beyond an introduction*, 2nd ed., Boston, MA: Birkhäuser, 2002.
- [Kod67] T. Kodama, *On the law of product in the Hecke ring for the symplectic group*, Mem. Fac. Sci., Kyushu Univ., Ser. A **21** (1967), 108–121.
- [Lan76] R. Langlands, *On the functional equations satisfied by Eisenstein series*, Lecture Notes in Mathematics 544. Springer-Verlag, 1976.
- [LV07] E. Lindenstrauss and A. Venkatesh, *Existence and Weyl’s law for spherical cusp forms*, Geom. Funct. Anal. **17** (2007), no. 1, 220–251.
- [Mar] S. Marshall, *L^p norms of higher rank eigenfunctions and bounds for spherical functions*, arXiv:1106.0534
- [Mil10] D. Milićević, *Large values of eigenfunctions on arithmetic hyperbolic surfaces*, Duke Math. J. **155** (2010), no. 2, 365–401.
- [Mil11] ———, *Large values of eigenfunctions on arithmetic hyperbolic 3-manifolds*, Geom. Funct. Anal. **21** (2011), no. 6, 1375–1418.
- [Nzo83] B. Nzoukoudi, *Représentations irréductibles unitaires de $Sp(2, \mathbb{R})$* , C. R. Acad. Sci., Paris, Sér. I **297** (1983), 451–454.
- [PS09] A. Pitale and R. Schmidt, *Ramanujan-type results for Siegel cusp forms of degree 2*, J. Ramanujan Math. Soc. **24** (2009), 87–111.
- [RoSc07] B. Roberts and R. Schmidt, *Local newforms for $GSp(4)$* , LNM 1918, Springer-Verlag 2007
- [RuSa94] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Commun. Math. Phys. **161** (1994), no. 1, 195–213.
- [RySh08] N. Ryan and T. Shemanske, *Inverting the Satake map for Sp_n and applications to Hecke operators*, Ramanujan J. **17** (2008), no. 2, 219–244.
- [Sar] P. Sarnak, *Letter to Morawetz*, available at <http://www.math.princeton.edu/sarnak>.
- [Sar11] ———, *Recent progress on the quantum unique ergodicity conjecture*, Bull. Am. Math. Soc., New Ser. **48** (2011), no. 2, 211–228.
- [SS89] A. Seeger and C. Sogge, *Bounds for eigenfunctions of differential operators*, Indiana Univ. Math. J. **38** (1989), no. 3, 669–682.
- [Sel56] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc., New Ser. **20** (1956), 47–87.
- [Shi63] G. Shimura, *On modular correspondences for $Sp(n, \mathbb{Z})$ and their congruence relations*, Proc. Nat. Acad. Sci. U.S.A. **49** (1963), 824–828.
- [SV] L. Silberman and A. Venkatesh, *Entropy bounds for Hecke eigenfunctions on division algebras*, preprint.
- [SZ02] C. Sogge and S. Zelditch, *Riemannian manifolds with maximal eigenfunction growth*, Duke Math. J. **114** (2002), no. 3, 387–437.
- [Ste93] E. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series, vol. 43, Princeton University Press, Princeton, NJ, 1993.
- [Tem] N. Templier, *Large values of modular forms*, Cambridge J. Math., to appear (arXiv:1207.6134)
- [Van97] J. VanderKam, *L^∞ norms and quantum ergodicity on the sphere*, Int. Math. Res. Not. **1997** (1997), no. 7, 329–347.
- [Ven10] A. Venkatesh, *Sparse equidistribution problems, period bounds and subconvexity*, Ann. Math. (2) **172** (2010), no. 2, 989–1094.
- [Zel10] S. Zelditch, *Recent developments in mathematical quantum chaos*, Current developments in mathematics, 2009, Somerville, MA: International Press, 2010, pp. 115–204.

MATHEMATISCHES INSTITUT, BUNSENSTR. 3-5, 37073 GÖTTINGEN, GERMANY

E-mail address: blomer@uni-math.gwdg.de

E-mail address: pohl@uni-math.gwdg.de